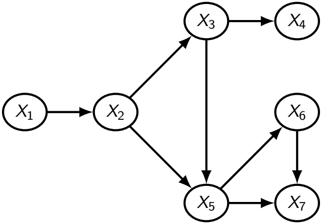
A polynomial algorithm to compute exact derivatives of the likelihood in Bayesian networks

Alexandra Lefebvre and Grégory Nuel
(PhD supervisors: A. Duval and G. Nuel)

LPSM, CNRS 8001, *Sorbonne Université, Paris, France*

Journées Francophones sur les Réseaux Bayésiens
et les Modèles Graphiques Probabilistes
Toulouse, 2018

Context



$X_{\mathcal{U}} = \{X_u\}_{u \in \mathcal{U}}$
pa_u : set of labels for parents of X_u
$\mathcal{X}_u$: set of values for X_u
$\mathrm{ev} = \{X_u \in \mathcal{X}_u^* \subset \mathcal{X}_u\}, u \in \mathcal{U}$
$K_u(X_u | X_{\mathrm{pa}_u}; \theta) = \mathbb{P}(X_u, X_u \in \mathcal{X}_u^* | X_{\mathrm{pa}_u}; \theta)$

$$L(\theta) = \mathbb{P}(\mathrm{ev}|\theta) = \sum_{X_{\mathcal{U}}} \prod_{u \in \mathcal{U}} K_u(X_u | X_{\mathrm{pa}_u}; \theta)$$

Our aim :

Compute $L^{(k)}(\theta)$ for $k = 0, 1, 2, \ldots$

State of art and our objective

- **Sensitivity analysis**: express $L(\theta)$ as a polynomial in θ
  $\Rightarrow$ not always possible (*e.g.* $\mathbb{P}(X_u = 1) = e^\theta/(1 + e^\theta)$)
  $\Rightarrow$ prohibitive if degree in θ is large (*e.g.* θ in many potentials)
- **Smoothing recursions** (Cappé and Moulines, 2005): express $\nabla^d \log L(\theta)$ as function of $\mathbb{E}[\nabla^d \log K_u|\mathrm{ev}]$ (Fisher-Louis' formulas)
  $\Rightarrow$ developed for Hidden Markov Models
  $\Rightarrow$ many terms and complex recursions when d increases
- **Polynomial computations**: perform sum-product with polynomials
  $\Rightarrow$ moments of additive functional in BN (Cowel, 1992; Nilsson, 2001)
  $\Rightarrow$ mgf/pgf for regular expr. in Markov models (Nuel, 2008, 2010)

Our idea: Use polynomial computations in the framework of derivatives

Method

- Definition : Derivative generating function (Dgf)
  Let f be a $\mathcal{C}^d$ function of θ, $d \in \mathbb{N}$

$$D^d f(\theta) = \sum_{k=0}^{d} f^{(k)}(\theta) z^k$$

  Example: $D^2 f(\theta) = f(\theta) + f'(\theta)z + f''(\theta)z^2$

- Polynomial potential of degree d :

$$\forall u \in \mathcal{U}; \quad D^d K_u(X_u | X_{\mathrm{pa}_u;\theta}) = \sum_{k=0}^{d} {K_u}^{(k)}(X_u | X_{\mathrm{pa}_u;\theta}) z^k$$

- "Leibniz product" $\bigstar$

$$L(\theta) = \sum_{X_{\mathcal{U}}} \prod_{u \in \mathcal{U}} K_u(X_u | X_{\mathrm{pa}_u}; \theta) \Rightarrow D^d L(\theta) = \sum_{X_{\mathcal{U}}} \bigstar_{u \in \mathcal{U}} D^d K_u(X_u | X_{\mathrm{pa}_u}; \theta)$$

“Leibniz product” : ★

Définition “Leibniz product”

$$P = \sum_{k=0}^{d} a_k z^k \quad \text{with} \quad a_k = f^{(k)}(\theta)$$

$$Q = \sum_{k=0}^{d} b_k z^k \quad \text{with} \quad b_k = g^{(k)}(\theta)$$

$$P \star Q = \sum_{k=0}^{d} c_k z^k \quad \text{with} \quad c_k = (fg)^{(k)}(\theta)$$

$$(fg)^{(k)}(\theta) = \sum_{i=0}^{k} \binom{k}{i} f^{(i)}(\theta)\, g^{(k-i)}(\theta) \Rightarrow c_k = \sum_{i=0}^{k} \binom{k}{i} a_i b_{k-i}$$

$$P \star Q = \sum_{k=0}^{d} \sum_{i=0}^{k} \binom{k}{i} a_i b_{k-i} z^k$$

Multidimensional parameter $\Theta = \theta_1, \ldots, \theta_p$

- Dgf of f function $\mathcal{C}^d$ of $\Theta \in \mathbb{R}^p$

$$D^d f(\Theta) = \sum_{k_1+\ldots+k_p \leq d} \frac{\partial^{(k_1+\ldots+k_p)} f(\theta)}{\partial {\theta_1}^{k_1} \ldots \partial {\theta_p}^{k_p}} {z_1}^{k_1} \ldots {z_p}^{k_p}$$

- "Leibniz product"

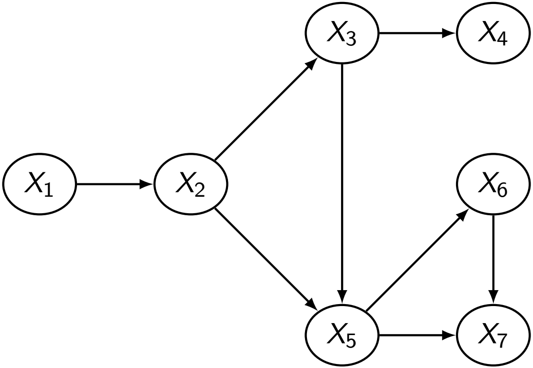
$$P = \sum_{k_1+\ldots+k_p = d} a_{k_1,\ldots,k_p} {z_1}^{k_1} \ldots {z_p}^{k_p} \; ; \; a_{k_1,\ldots,k_p} = \frac{\partial^{(k_1+\ldots+k_p)} f(\Theta)}{\partial {\theta_1}^{(k_1)} \ldots \partial {\theta_p}^{(k_p)}}$$

Q idem with coef. b instead of a and function g instead of f

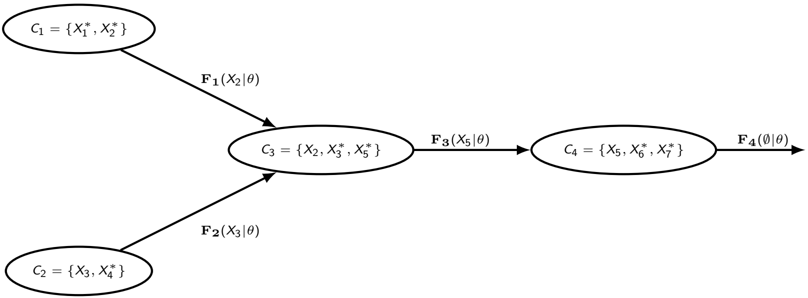
$$P \star Q = \sum_{k_1+\ldots+k_p = d} c_{k_1,\ldots,k_p} {z_1}^{k_1} \ldots {z_p}^{k_p}$$

with $$c_{k_1,\ldots,k_p} = \sum_{i_1=0}^{k_1} \ldots \sum_{i_p=0}^{k_p} \binom{k_1}{i_1} \ldots \binom{k_p}{i_p} a_{k_1-i_1,\ldots,k_-i_p} b_{i_1,\ldots,i_p}$$

Computation : $\sum \prod K_u \to \sum \bigstar D^d K_u$



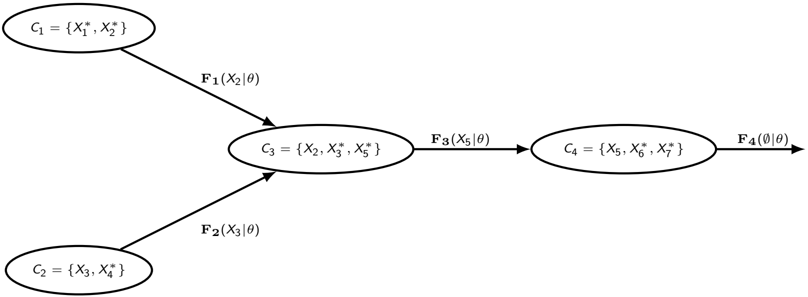
Computation : $\sum \prod K_u \to \sum \bigstar D^d K_u$



$$\forall i \in \{1, \dots 4\}; \quad \mathbf{\Phi}_i^d(C_i|\theta) = \bigstar_{X_u \in C_i{}^*} D^d K_u(X_u|X_{\mathrm{pa}_u};\theta)$$

$$\boldsymbol{F}_i^d(S_i|\theta) = \sum_{C_i \backslash S_i} \left(\bigstar_{j \in \mathrm{from}_i} \boldsymbol{F}_j^d(S_j|\theta) \right) \star \mathbf{\Phi}_i^d(C_i|\theta)$$

Computation : $\sum \prod K_u \to \sum \bigstar D^d K_u$



$$\mathbf{\Phi}_1^d(X_1, X_2|\theta) = D^d K_1(X_1; \theta) \star D^d K_2(X_2|X_1; \theta)$$

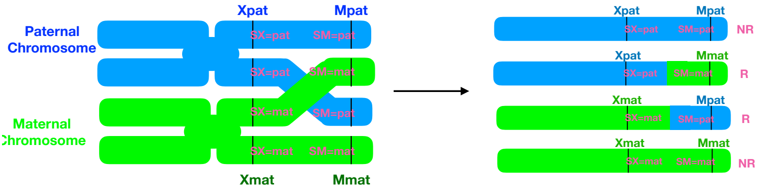
$$\boldsymbol{F}_3^d(X_5) = \sum_{X_2} \sum_{X_3} \boldsymbol{F}_1^d(X_2|\theta) \star \boldsymbol{F}_2^d(X_3|\theta) \star \mathbf{\Phi}_3^d(X_2, X_3, X_5|\theta)$$

$$\boldsymbol{F}_4^d(\emptyset|\theta) = D^d L(\theta) = \sum_{k=0}^{d} L^{(k)}(\theta) z^k$$

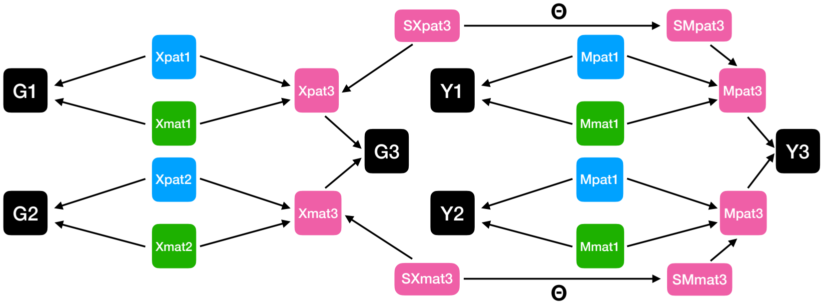
Application : two-point linkage in genetics

- Goal : localize a targeted gene X driving phenotype Y
- Based on pedigrees (family structure)
- Observed :
  - Genotype $G = \{M_{\text{pat}}, M_{\text{mat}}\}$ of marker with known localization
  - Phenotype of trait Y, known penetrance $\mathbb{P}(Y|X_{\text{pat}}, X_{\text{mat}})$
- Latent X_{pat}, X_{mat}, $SX_{\text{pat}}, SX_{\text{mat}}, SM_{\text{pat}}, SM_{\text{mat}}$
- linkage = correlation between M and X :

$$\theta = \frac{\#\text{R}}{\#\text{R} + \#\text{NR}} \qquad \text{distance} = -\frac{1}{2}\log(1 - 2\theta)$$



DAG for two-point linkage



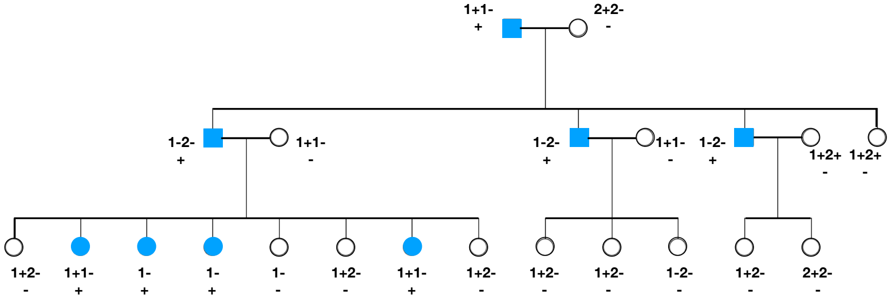
$$D^2\mathbb{P}(SX_{\text{pat}_3} = \text{pat}) = D^d\mathbb{P}(SX_{\text{pat}_3} = \text{mat}) = 0.5 \text{ (Mendelian transm.)}$$

$$D^2\mathbb{P}(SM_{\text{pat}_3} = SX_{\text{pat}_3} | SX_{\text{pat}_3}; \theta) = (1 - \theta) - z$$

$$D^2\mathbb{P}(SM_{\text{pat}_3} = SX_{\text{pat}_3} | SX_{\text{pat}_3}; \beta) = \frac{1}{1 + e^\beta} - \frac{e^\beta}{(1 + e^\beta)^2} z + \frac{e^\beta(e^\beta - 1)}{(1 + e^\beta)^3} z^2$$

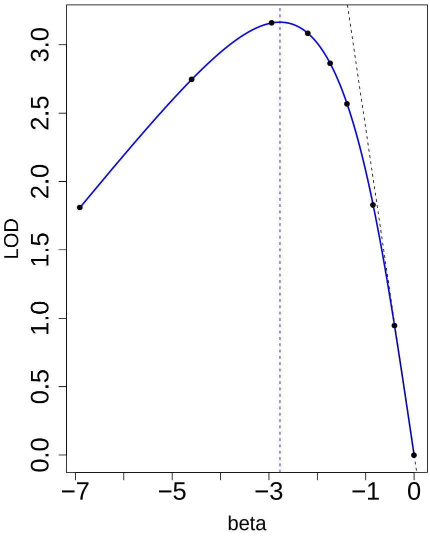
Application with the two-point linkage

Ped2a example in Mendel package (First family : KUS)



Marker: PGM1 locus, allele $M \in \{1+, 1-, 2+, 2-\}$, $q_M = (0.620, 0.170, 0.140, 0.070)$

Trait: RADIN locus, allele $X \in \{-, +\}$, $q_X = (0.997, 0.003)$, $+$ dominant, NB: genotype $+/+$ considered lethal.



$$\theta = \frac{e^{\beta}}{1+e^{\beta}} \text{ ie } \beta = \log\left(\frac{\theta}{1-\theta}\right)$$

$$\mathrm{LOD}(\beta) = \frac{\ell\left(\frac{e^{\beta}}{1+e^{\beta}}\right) - \ell(0.5)}{\log(10)}$$

$$\hat{\beta} = -2.772409$$

$$\mathrm{LOD}(\hat{\beta}) = 3.164775$$

$$\frac{\partial \mathrm{LOD}}{\partial \beta}(0.0) = -2.38862$$

$$\frac{\partial^2 \mathrm{LOD}}{\partial \beta^2}(\hat{\beta}) = -0.4104178$$

Black dots computed with Mendel 16.0

Summary and perspectives

Summary:

- Sum-product with polynomials and "Leibniz product"
- univariate: derivatives up to order d in $\mathcal{O}(d^2 \times C)$
- multivariate $\theta \in \mathbb{R}^p$: gradient and Hessian in $\mathcal{O}(p^2 \times C)$
- confidence intervals, theoretical distributions under H0 and H1

Perspectives:

- score and Wald tests for two-point linkage
- joint estimation of allele frequencies and θ
- extensive simulations under H0 and H1
- other genetic models: segregation, TDT, IBD, etc.